
Тезаурусы и Онтологии в научно-образовательных информационных системах.

Одним из основных результатов созидательной, социальной и интеллектуальной человеческой деятельности является создание и накопление информационных ресурсов с целью их дальнейшего использования и недопущения утраты опыта предыдущих поколений.

*Окружающий нас мир
непознаваем, ввиду того, что
мы изучаем не его, а лишь
наше представление о нем*

Эммануил Кант

Не будет преувеличением сказать, что уровень развития технологий накопления информации и эффективности использования накопленной ранее информации на протяжении всей истории человечества значительно влиял на уровень развития производительных сил. Утеря информации приводила к отбрасыванию цивилизации на века назад.

Однако, чтобы эффективно пользоваться накопленной ранее информацией, необходимы специальные инструменты и специальные технологии, при помощи которых могут быть реализованы специальные приемы работы с информацией.

*Многие вещи нам не
понятны не потому, что
наши понятия слабы: но
потому, что сии вещи
не входят в круг наших
понятий*

Козьма Прутков

*Если Вам все равно, где Вы находитесь
значит Вы не заблудились*

Одна из вечных проблем человеческого сообщества. На протяжении своего многотысячелетнего развития его представители неустанно находятся в поиске того, где находится что-либо: *пищи, жилища, пастбищ, дорог, сокровищ* и т. п.

Человечество постоянно находится в поиске знаний: «информации о том, где лежат сокровища».

Великий аргентинский писатель Хорхе Луис Борхес в своем эссе «Четыре цикла» писал, что в мировой литературе вечными являются четыре темы:

- Падение города
- Возвращение героя
- Поиск
- Самопожертвование бога.



«Есть правила для выбора решения, но нет правил для выбора этих правил»

Энон

- Нетрудно заметить, что наиболее часто встречающейся как в литературе, так и в реальности является третья тема - **поиск**, ибо четвертая тема выходит за рамки обычного человеческого опыта, а две первые проявляются лишь в «минуты мира роковые»

- «Вавилонская библиотека» притча 1941

В этой притче Вселенная представляется в виде Библиотеки, беспредельной и всеобъемлющей, на полках которой можно обнаружить все возможные комбинации двадцати с чем-то орфографических знаков (число их, хотя и огромно, но не бесконечно) или все, что поддается выражению — на всех языках.. Философский смысл притчи, конечно же, гораздо глубже проблемы информационного поиска, но исходный образ взят автором из повседневной реальности

Когда было провозглашено, что Библиотека объемлет все книги, первым ощущением была безудержная радость. Каждый чувствовал себя владельцем тайного и нетронутого сокровища.

Не было проблемы — личной или мировой, для которой не нашлось бы убедительного решения . .

. Вселенная обрела смысл, вселенная стала внезапно огромной, как надежда. В это время много говорилось об Оправданиях: книгах апологии и пророчеств, которые навсегда оправдывали деяния каждого человека во вселенной и хранили чудесные тайны его будущего. Тысячи жаждущих покинули родные шестигранники и устремились вверх по лестницам, гонимые напрасным желанием найти свое оправдание . . . , но те, кто пустился на поиски, забыли, что для человека вероятность найти свое Оправдание или какой-то его искаженный вариант равна нулю . . . На смену надеждам, естественно, пришло безысходное отчаяние.

Мысль, что на какой-то полке в каком-то шестиграннике скрываются драгоценные книги и что эти книги недосыгаемы, оказалась почти невыносимой. Одна богохульная секта призывала всех бросить поиски и заняться перетасовкой букв и знаков, пока не создадутся благодаря невероятной случайности канонические книги. . .

На некоей полке в некоем шестиграннике (полагали люди) стоит книга, содержащая суть и краткое изложение всех остальных: некий библиотекарь прочел ее и стал подобен Богу. В языке этих мест можно заметить следы культа этого работника отдаленных времен. Многие предпринимали паломничество с целью найти Его. В течение века шли безрезультатные поиски. Как определить таинственный священный шестигранник, в котором Он обитает? Кем-то был предложен регрессивный метод: чтобы обнаружить книгу А, следует предварительно обратиться к книге В, которая укажет место А; чтобы разыскать книгу В, следует предварительно справиться в книге С, и так до бесконечности

- Проблема поиска информации - одна из вечных проблем человечества. Чтобы решить проблему доступа к информации, человечество создало библиотеки - универсальную систему хранения, систематизации и каталогизации «информации и знаний».
- Александрийская библиотека

- В процессе научной, а особенно образовательной, деятельности много времени и сил отнимает работа с литературными источниками, разного рода материалами и документами: поиск необходимых документов, систематизация и классификация документов в соответствии с поставленной задачей.
- Для удовлетворения информационных потребностей современных пользователей необходима поддержка сложных функций поиска и классификации информации, а также просмотр ресурсов по категориям (рубрикам) и словарям-классификаторам.

- Наиболее важной задачей является задача систематизации ресурсов, для решения которой необходимо четко определить состав логико-семантических категорий (фасетов) и ключевых терминов (понятий), покрывающих избранную достаточно узкую предметную область, интересующую пользователя.
- Как правило предметная область ограничивается изучаемым учебным курсом или конкретной темой курса.

- Стандартным подходом к систематизации информации является классификация документов с помощью таксономий.
- Таксономия – это предметная классификация, которая группирует термины в виде управляемого словаря (тезауруса) и упорядочивает их (словари) в виде иерархических структур.
- Основу классификации составляет выделение понятий (ключевых терминов), установление парадигматических отношений (например, типа родитель – потомок) между ними и сопоставление анализируемого документа выделенным ПОНЯТИЯМ.

- Для описания какой-либо предметной области обычно используется определенный набор ключевых терминов, каждый из которых обозначает или описывает какое-либо понятие из данной предметной области.



Термины в области лексики и формула в области синтаксиса являются теми идеальными типами языкового выражения, к которым неизбежно стремится научный язык.

Шарль Балли

- По определению международной организации по стандартизации (ISO), тезаурус является словарем, управляемым языком индексации, формально организованным для того, чтобы установить явные априорные отношения между понятиями.

- Индексация (Индексирование) — это сортировка (систематизация) информации по каким-либо формальным принципам. Различают два вида индексирования: классификационное и координатное.

- Это определение устанавливает лексические единицы (термины) и семантические отношения между этими единицами в качестве элементов, составляющих тезаурус.
- Тезаурусные (**парадигматические**) отношения (род-вид, часть-целое, комплекс-элемент, причина-следствие) налагаются на структуру таксономии, то есть идентифицируются основные таксономии предметной области.

- Тезаурус это словарь (идеографический или семантический словарь), в котором слова (лексические единицы, термины, понятия, дескрипторы) размещаются их по смысловой близости.

- Словари, в которых лексика располагается не алфавитном порядке, а на основании этого критерия, получили название **идеографических** (от греч. *idea* — понятие, идея, образ и *grapho* — пишу) или семантических (от др.-греч. *семантика* (*σημαντικός* — обозначающий) — раздел лингвистики, изучающий смысловое значение единиц языка).

- Тезаурус является инструментом концептуального (понятийного) описания отдельных предметных областей.
- В отличие от толкового словаря, тезаурус определяет понятия (ключевые термины) не только с помощью описания, данного в словарной статье, но и посредством соотнесения понятия с другими понятиями и их группами, благодаря чему может использоваться как система знаний, отраженных языком (словами), и, как было отмечено Шрейдером Ю. А., тогда тезаурус становится интересным сам по себе, а не только как вспомогательный инструмент.

- **Синонимия** — связь между терминами, различными по написанию, но имеющих одинаковое или очень близкое понятийное (лексическое) значение, например: кавалерия – конница, смелый – храбрый.
- **Антонимия** — связь между терминами, различными по написанию, имеющих прямо противоположные значения: правда – ложь, добрый – злой.

- **Гипонимия/гиперонимия** *Гипероним* — термин (понятие, слово) с более широким значением, выражающее общее, родовое понятие, название класса (множества) предметов (свойств, признаков). *Гипоним* — термин с более узким значением, называющее предмет (свойство, признак) как элемент класса (множества). Эти отношения транзитивны и несимметричны. Гипоним наследует все свойства гиперонима.
- **Меронимия/партонимия** — отношение типа «часть-целое», с выделенными отношениями «быть элементом» и «быть сделанным из».

- Для описания какой-либо предметной области всегда используется определенный набор терминов, каждый из которых обозначает или описывает какое-либо понятие или концепцию из данной предметной области. Совокупность терминов, описывающих данную предметную область, с указанием семантических отношений (связей) между ними является *тезаурусом*. Такие отношения в тезаурусе всегда указывают на наличие смысловой (семантической) связи между терминами.
- Основным отношением (связью) между терминами в тезаурусе является связь между *более широкими* (более выразительными) и *более узкими* (более специализированными) понятиями. Часто выделяют 2 подвида этого отношения:
 - Один термин обозначает понятие, являющееся частью понятия, обозначаемого другим термином (например, «наука» и «математика», «математика» и «теория чисел»)
 - Один термин обозначает элемент класса, обозначаемого другим термином («горные районы» и «Кавказ»).

- Существуют также и другие связи между терминами. Например, одно понятие или концепция может быть обозначено несколькими терминами, являющимися синонимами. Некоторые термины могут быть антонимами для других. Часто среди терминов, относящихся к одному понятию, выделяют единственный (для каждого языка тезауруса) *наиболее предпочтительный (наиболее подходящий) термин*, который наиболее хорошо характеризует, или обозначает данное понятие. Остальные термины являются *менее предпочтительными (менее подходящими)*.
- Помимо вышеописанных, между терминами могут существовать также и другие, *ассоциативные связи*, если понятия, обозначаемые этими терминами, как-либо связаны между собою по своему смыслу, за исключением описанных выше иерархических связей.

- Термины могут иметь следующие атрибуты (основные):
 - **ID** – *Identifier*. Уникальный идентификатор термина.
 - **SN** – *Scope Note*. Комментарий к термину. Например, представляет вербальное пояснение термина, или правила его использования.
 - **TT** – *Top Term*. Признак, Выделяющий термины на самом верхнем уровне иерархии (термины наиболее общих понятий в данной иерархии понятий).
 - **HN** – *History Note*. История модификации связей и атрибутов данного термина.

- Термины могут иметь следующие отношения (основные):
 - **USE** – Связывает термин с наиболее предпочтительным (на том же языке) термином для данного понятия.
 - **UF** – *Used For*. Обращение связи **USE**. Связывает наиболее подходящий термин с синонимами и квазисинонимами (менее подходящими терминами).
 - **BT** – *Broader Term*. Связь термина с термином более общего понятия.
 - **NT** – *Narrower Term*. Обращение связи **BT**.
 - **RT** – *Related Term*. Ассоциативная связь. Связывает семантически связанные между собою термины, не находящиеся при этом в одной иерархии, и не являющиеся синонимами или квазисинонимами.
 - **LE** – лексический эквивалент термина (на другом языке)

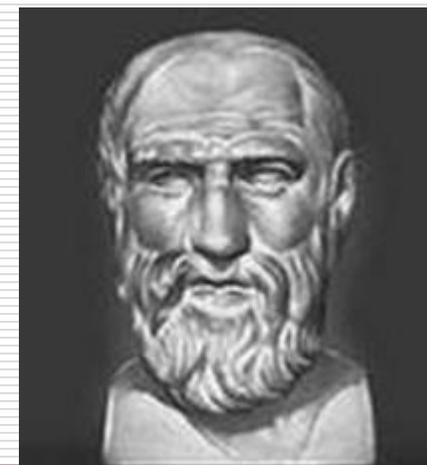
- Необходимость в расположении слов по сходству, смежности, аналогии их значений ощущалась на всем протяжении обозримой истории человеческой мысли.
- На заре цивилизации, когда люди могли выразить свои мысли на письме лишь при помощи идеограмм и символов, единственно возможным словарем был, вероятно, такой, в котором слова располагались по тематическим группам.
- Лексикографу в то время просто трудно было найти иной критерий для классификации слов, кроме отношений, существующих в самой действительности.

- История возникновения тезаурусов восходит к великим мыслителям древности и, прежде всего, к Аристотелю.
- Ему принадлежат слова, возраст которых 2,5 тыс. лет: «Из слов, высказываемых без какой-либо связи, каждое означает или сущность, или качество, или количество, или отношение, или обладание, или действие, или страдание».



Тезаурус – это словарь, в котором слова и словосочетания с близкими значениями сгруппированы в единицы, называемые понятиями, концептами или дескрипторами, и в котором явно (в виде отношений, иерархии) указываются семантические отношения между этими понятиями (концептами, дескрипторами).

- Одной из наиболее древних попыток идеографических классификаций является труд Аристофана Византийского (с 195 г. до н.э. директор Александрийской библиотеки, умер в 180 г. до н.э.): «Аттический глоссарий»



- В I веке н. э. появляется «Словарь синонимов» грекоязычного финикийского ученого Филона Библского, который называют одним из первых тезаурусов.
- Филон Библский (Геренний Филон, лат. Herennius Philon, ок. 50 — после 138 гг. н.э.) — финикийский историк, грамматик, родом из города Библа. Он составил на греческом языке труд по истории и мифологии Финикии, «О городах и о том, что замечательного в них было» в 30 книгах, «О приобретении и отборе книг» в 12 книгах, «Словарь синонимов», «Об иудеях».

- Во II в. н. э. появляется капитальный труд на материале греческого языка, составленный лексикографом и софистом Юлием Поллуксом (настоящее имя Полидевк – Юлий Полидэвк, др.греч. Ιούλιος Πολυδεύκης), уроженцем египетского города Навкратис.



- Ю. Поллукс, пользовавшийся покровительством римских императоров Марка Аврелия и его сына Коммода. Преподавал риторику в Афинах, где и оставался до конца жизни.
- Написал несколько сочинений, из которых до нас дошел лишь его словарь «Ономастикон» («Искусство наименования»), причём не в оригинальном варианте, а в более позднем переложении.

- Словарь Ономастикон состоит из 10 книг. Каждая из них начинается приветственным письмом к императору Коммоду. Книги по существу являются отдельными трактатами и содержат в себе наиболее важные слова, относящиеся к той или иной теме. Так, в первой книге говорится о богах и царях, во второй — о людях, их жизни и физиологическом строении, в третьей — о родстве и гражданских отношениях и т. д.

- Слова, помещенные в словарь, сопровождаются краткими толкованиями.
- В новое время словарь был впервые опубликован в 1502 г. в Венеции.

- Словарь составлен по предметному принципу, с включениями элементов риторики, содержит краткие толкования слов, множество цитат из более ранних авторов: Аристофана Византийского, Эратосфена, Памфила и других, сведения из разных областей жизни, а также 52 выражения, служащие для восхваления правителей, и 33 выражения для ругательств в адрес сборщиков податей.

- Боги. Места культа, алтари и храмы. Создание и разрушение. Жрецы. Провидцы и искусство провидения. Благочестивые и безбожники. Короли, купцы, ремесленники. Дом, корабль. Погода. Армия. Лошади и искусство верховой езды. Домашние животные. Сельское хозяйство, плуг, средства перемещения, пчелы.
- Человек. Возрастные отличия. Рождение человека. Части тела.
- Пол. Родство. Брак. Дети. Друзья. Господа и рабы. Строительное дело. Географическое положение. Путешествие. Печаль, радость.
- Образование: грамматика и риторика. Философы и софисты. Поэты и музыканты. Музыкальные инструменты. Танцы, театр. Астрономия. Медицина и болезни.
- Охота. Собаки. Животные, на которых охотятся. Женские украшения. Мужество, страх. Фармацевтическое дело. Молитва. Слава.
- Гости. Вино и продукты. Еда. Застольная беседа.
- Рынок. Купля и продажа. Торговцы. Товары. Деньги.
- Суд. Судья. Процессы. Наказания. Доносчик.
- Административное деление. Город. Общественные здания. Игры детей и взрослых.
- Утварь.

- Между II и III веком н. э. появляется санскритский словарь «Амаракоша» (санскр. अमरकोशः Amarakośa -- Коша -- сокровище Амара), другое название «Намалинганушасана» (санскр. नामलिङ्गानुशासनम्: Nāmaṅgānuśāsanam), от «нама-линга-ану-шасана» — наставление о существительных и грамматических родáх) который содержал около 10 тыс. слов и состоял из трех книг, каждая из которых делилась на главы, главы — на секции.

- Так, первая книга была посвящена небу, богам и всему тому, что с ними связано. В ней имелись секции «времена года», «небесный свод» и т.п.
- Вторая книга содержала слова, относящиеся к земле, растительному и животному миру и человеку.
- Для лучшего запоминания толкования давались в стихотворной форме.

- Его автором является древнеиндийский поэт, грамматик и лексикограф Амара Сина (по другим источникам Амара Дэва), которого называли «одной из девяти жемчужин, украшающих трон Викрамадितье».
- Чандрагупте II Викрамадितье --- один из царей государства Гупт в Древней Индии.

- Словарь составлен в строгом соответствии с уровнем развития науки и господствующими представлениями в обществе, что позволяет нам судить о понимании и объяснении мира, об устройстве человеческого общества в те далекие времена.

- Словарь Амаракоша разительно отличается от алфавитных словарей. Так в алфавитных словарях порядок записей (терминов) регулируется алфавитом, являющимся в значительной мере нейтральным инструментом, то при построении идеографического словаря решающее значение приобретает мировоззрение самого лексикографа. Даже тот факт, что автор словаря был буддистом нашел отражение в соответствующем разделе словаря.

«Амаракоша»

КНИГА I□ **Глава 1. Секции:**

Небо, боги, титаны и их атрибуты.

Небесный свод, атмосфера, планеты, звезды.

Времена года, фазы луны, затмения.

Преступление, добродетель, счастье, судьба, характер, ум, чувства, вкус, запах, цвет.

Слово, язык, сочинение.

Звук.

Музыка, танцы, драматические представления, празднества.

□ **Глава 2. Секции:**

Области ада, темь, змеи, яды.

Потусторонний мир, души усопших, несчастье, страдание.

Моря, вода, острова, реки, суда, рыба, водоемы, водоросли.

КНИГА II

□ **Глава 1.** Земля, солнце, страна, дороги, меры длины.

□ **Глава 2.** Города, здания, жилище.

□ **Глава 3.** Горы, скалы, источники, пещеры, минералы.

□ **Глава 4. Секции:**

■ Леса, сады, деревья, растения, части растения.

■ Деревья разных пород.

■ Лекарственные растения.

■ Полезные растения.

■ Огородные растения, травы.

□ **Глава 5.** Львы и другие четвероногие, насекомые, птицы, стаи, стада..

□ **Глава 6. Секции:**

■ Мужчины, женщины, родственники, государства, учреждения.

■ Здоровье, лекарства, болезни, части тела.

■ Одежда, украшения, благовония, гирлянды.

□ **Глава 7.** Расы, секты, режимы, сословие жрецов, жертвоприношение, милостыня, самоистязание, учение, брак, цель человеческой жизни.

□ **Глава 8. Секции:**

■ Сословие воинов, короли, министры, враги, союзники, оборона, победа, доходы, знаки королевского отличия.

■ Лагерь, армия, оружие, война, резня, похороны, тюрьма.

□ **Глава 9.** Третье сословие, профессии, земледельцы, поле, зерно, орудия труда, пища, тягловый скот, торговля, меры.

□ **Глава 10.** Четвертое сословие, арендаторы, ремесленники, артисты, музыканты, охота, охотники, собаки, дичь, орудия, умения, спиртные напитки, игрища.

- Термин «тезаурус» --- <<сокровище>>, аналог санскритского термина <<коша>>, был применен впервые в XIII веке в заголовке своего труда флорентийским поэтом и ученым Брунето Латини (итал. Brunetto Latini, ок. 1220—1294 - флорентийский поэт, учёный, государственный деятель) --- систематизированной энциклопедии, назвав ее «Книга о сокровище».



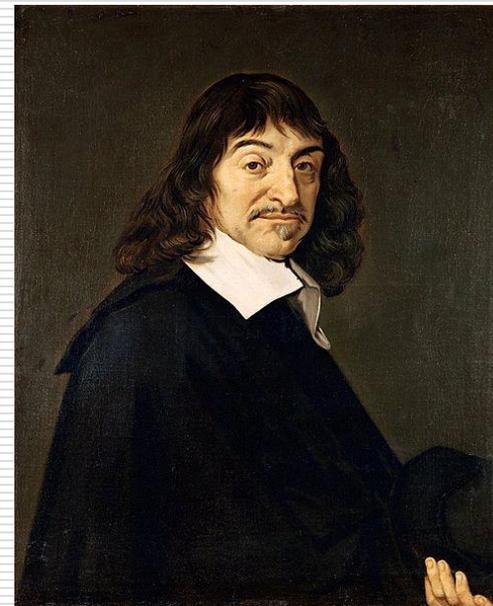
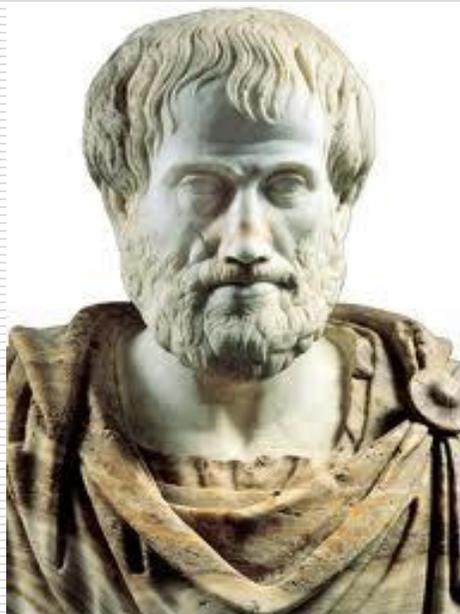
Ritratto di Brunetto Latini, nella Galleria di Firenze.

- В которой Брунетто даёт обзор знаний своего времени о Боге, природе, об истории древнего и нового времени, об искусстве; дает наставления для управления домом, государством и т.д. Это вполне соответствовало семантике употребленного слова «Thesaurus» (греч. *θησαυρος*, т.е. «сокровище», «богатство», «запас»).



Ritratto di Brunetto Latini, nella Galleria di Firenze.

Проблема формализации работы разума восходит к Аристотелю, а вопросы создания языка разума волновали Декарта.



Дальнейшее развитие идеи смысловой классификации лексики связано с проблемой так называемого всемирного философского языка. В основании этого языка должна лежать логическая классификация всего, что может быть предметом человеческой мысли. Значительный вклад в разработку этого вопроса внес испанский философ и богослов Раймонд Луллий {Раймонд Луллий (лат. *Raimundus Lullius*, исп. *Lulio*, кат. *Ramon Lull*, 1235--1315) --- каталанский поэт, философ и миссионер, один из наиболее ярких и оригинальных мыслителей европейского Средневековья..



Уже в XIII веке предпринимались попытки сформулировать процедуры поиска истинных высказываний.

Книга испанского монаха Раймонда Луллиля (1235-1315) "Великое и окончательное искусство" ("Ars Magna et Ultima") заметно повлияла как на современников, так и наше время.

Современным исследователям по искусственному интеллекту созвучны идеи многих мыслителей прошлого.

- В своем труде «Великое и окончательное искусство» ("Ars Magna et Ultima"), где он описал возможную конструкцию логической машины, Р. Луллий предлагает оригинальную систему, состоящую из 7 концентрических кругов. Каждый круг содержит обозначение группы понятий. Так, на одном приводятся девять субстанций: бог, ангел, небо, человек, воображаемое, чувственное, растительное, стихийное, инструментальное.

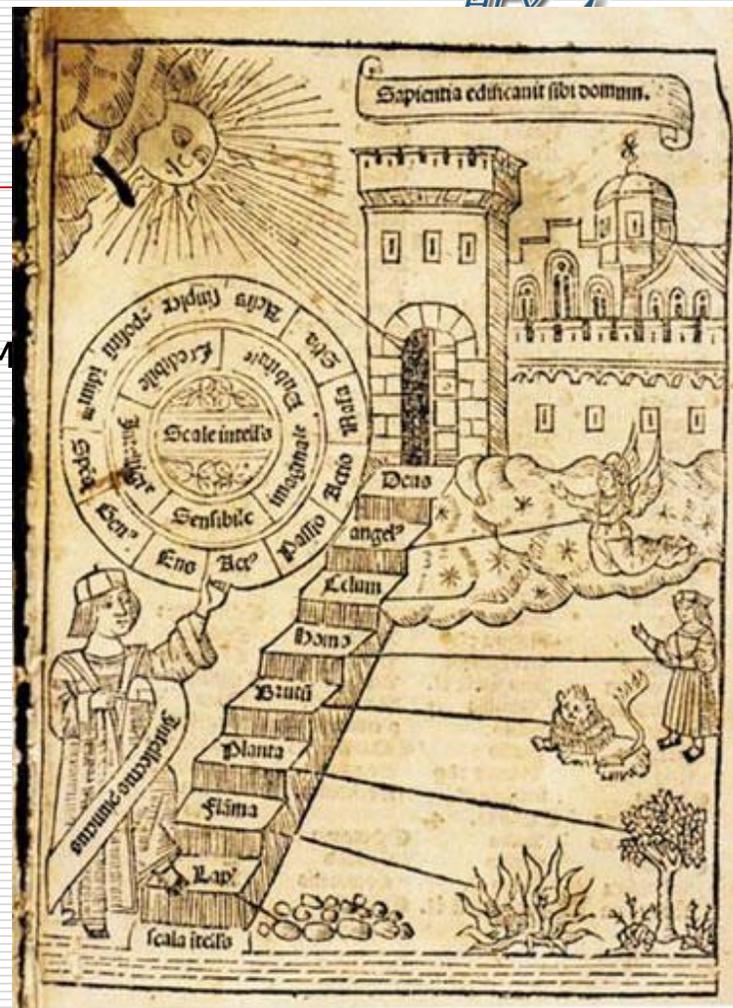
- Другой круг служит для указания на девять абсолютных предикатов: святость, величина, длительность, могущество, знание, стремление, добродетель, истина, слава; третий дает девять относительных предикатов: святой, великий, длительный, могучий, добродетельный и т. п.

- Кроме того, элементами системы испанского философа являются понятия, отражающие различные отношения между предметами (различие, тождество, противоречие, начало, середина, конец, больше, равно, меньше) и список вопросов (что?, чего?, почему?, как велико?, какого качества? и др.). При вращении кругов относительно друг друга возникают различные комбинации понятий, что дает возможность, по мнению Р. Луллия, находить истины и перечислять все предметы мысли.

- Одной из важных заслуг Р. Луллия у современников считалось подробное описание видов и сущностей хаоса — первоэлементов, которыми обозначались общие понятия или основные категории всего существующего, из которых создается весь мир, и отношений между ними, которые напоминают современные RDF-диаграммы.

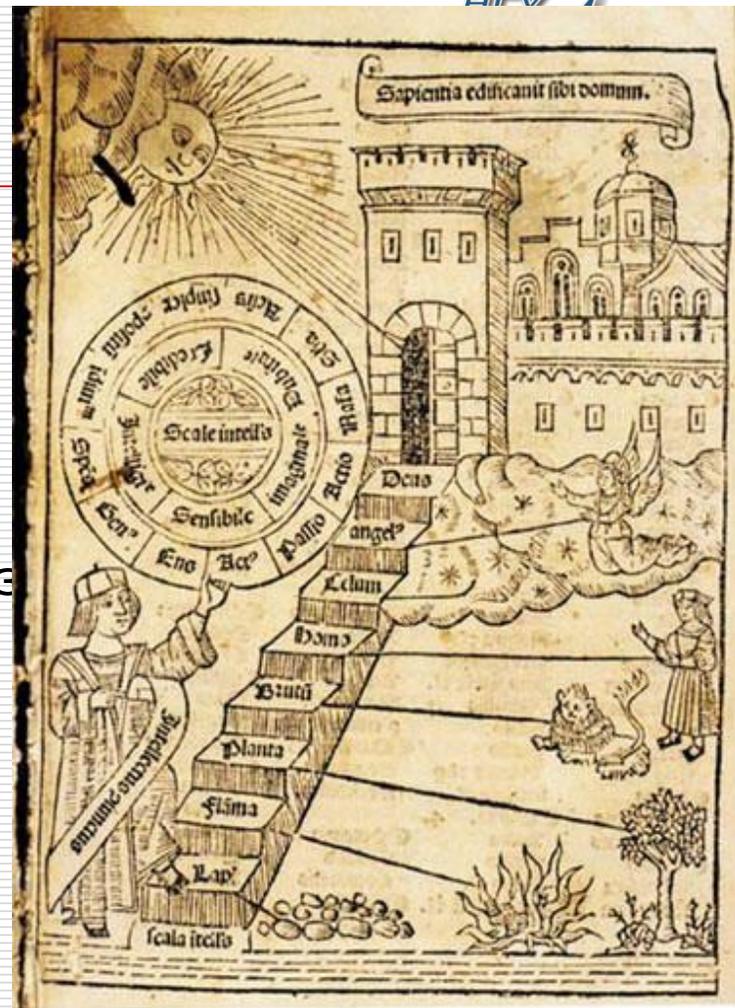
Искусство памяти Луллия (логическая машина) базируется на системе, посредством которой возможно формулировать множество теологических вопросов и ответов.

В нем используются символы, алфавит, обозначающий различные вещи, и потому кажется сходным с символической логикой. Сами буквы не наделены каким-либо значением, однако они выступают в роли определенных понятий, взятых из теологии и логики, которые и обозначают.



С другой стороны, Искусство обладает своего рода гибкостью, которая не была присуща **Аристотелевской логике**: та логика является словесной, она требует от оратора «вручную» подготавливать каждый из аргументов.

Кроме того, логика Аристотеля является синтаксической, а не семантической; то есть сказать, что «Все собаки - кошки; я - собака, следовательно, я - кошка», - возможно, но хотя утверждение является логичным, само заявление является ложным.



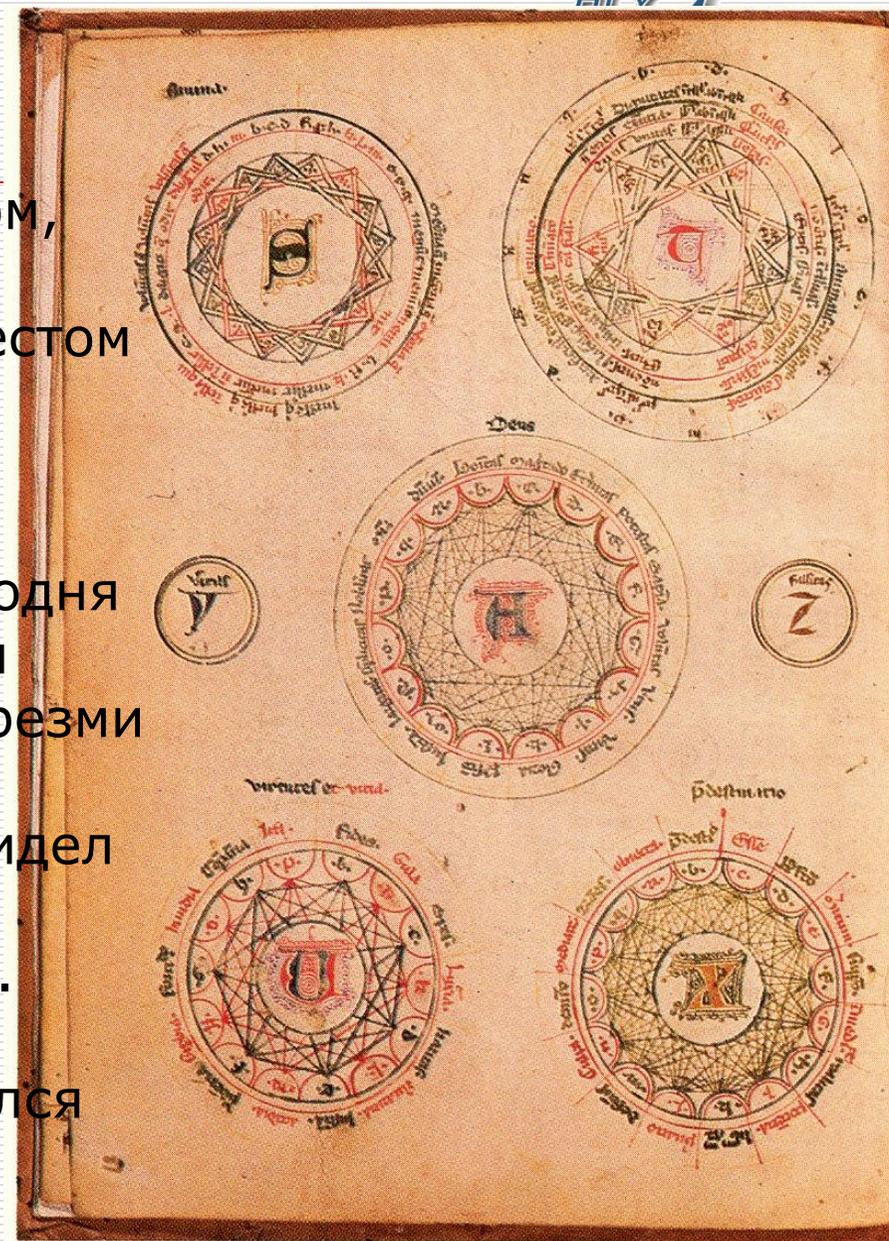
Многие математики рассматривают устройство Луллия как первую в истории попытку создать машину, способную выполнять ряд логических операций, осуществляемых человеческим мозгом. Луллий хотел свести процесс исследования реальности к механическому комбинированию небольшого количества исходных аксиом, отражающих отдельные фрагменты этой реальности.

Он считал, что в этих аксиомах неявно содержатся все истины науки, и что их можно извлечь из них путём чистого мышления, без обращения к опыту

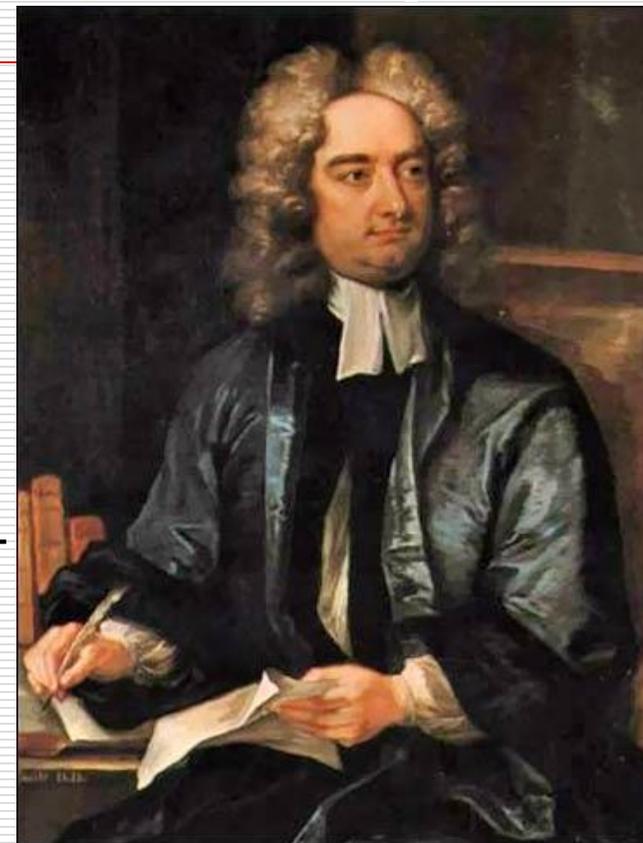


Поиск истины

Прежде всего, Луллий был миссионером, но, в отличие от инквизиторов, он, францисканец, хотел нести веру не крестом и мечом, а логикой и стремился быть доказательным в своей позиции. Он хотел говорить с иноверцами на их языке, используя для этого, как бы сегодня сказали, формальную логику. Ему были знакомы алгебраические труды Аль Хорезми и само понятие «алгоритм», а во время путешествий по Северной Африке он видел гадальные устройства, состоящие из концентрических вращающихся дисков. Сочетая алгоритмическую основу с дисковой механикой, Луллий намеревался создать инструмент для получения логических доказательств.



Джонатан Свифт (Jonathan Swift) — англо-ирландский писатель-сатирик, публицист, философ, поэт и общественный деятель. Наиболее известен как автор знаменитой фантастической тетралогии «Путешествия Гулливера», в которой остроумно высмеял человеческие и общественные пороки. В третьей части «Путешествий», где описан визит Гулливера в Великую академию, находящуюся в столице Лапутии, автор осмеивает некоторых не слишком прагматичных ученых. Специфическое отношение к ним Свифта можно понять, если учесть, что по своей «первой специальности» он был разведчиком. Не случайно его любили классики марксизма и советские литературоведы: уж им-то подобная критика науки точно пришлась по душе.



Переходя из комнаты в комнату, Гулливер в одной из них обнаруживает своеобразную конструкцию и выясняет, что огромная, площадью 20 квадратных метров рама представляет собой прибор для открытия отвлеченных истин. На раме располагаются таблички со словами, которые можно произвольно сочетать с помощью встроенного «генератора случайных чисел». Совершив очередную перетасовку, вовлеченные в эксперимент ученые пытаются найти в беспорядочном наборе слов и знаков осмысленные фразы и таким образом создать полный обзор всех наук и искусств. Эту машину Свифт назвал «компьютером», а прототипом ее создателя был Раймонд Луллий.



Свифт назвал машину Луллия «компьютером» вслед за Лейбницем, который считал, что научный спор можно решить с помощью бумаги, пера и вычислений. В английском переводе тезис Лейбница звучит так:

Gentlemen, let us compute!

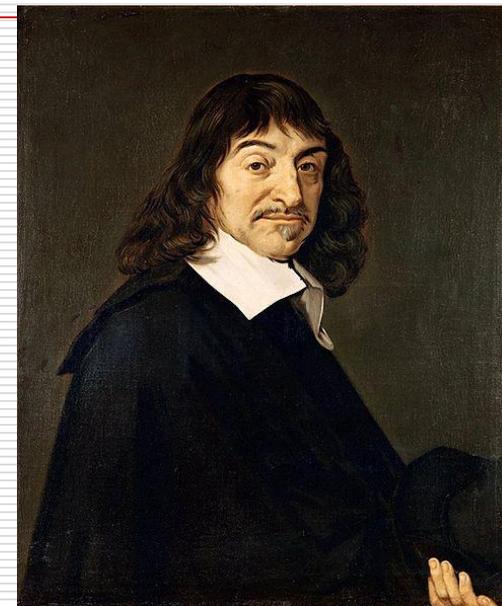
Надо иметь в виду, что до недавнего времени под словом compute понимали как доказательство чего-то математическими средствами, а не просто как вычисление.



Особое влияние идеи Луллия оказали на формирование философских и научных воззрений Г. В. Лейбница¹⁹. Он писал о возможности создания всеобщего философского языка, в основании которого лежала бы некоторая система понятий, и даже придумал название для этого языка (*Speciesus generalis*).

Так, Лейбниц заявлял, что его цель – не просто получение новых результатов (например, в математике), а выработка общего формального метода, позволяющего находить таковые. В основе, по Лейбницу, должен был лежать «всеобщий алфавит» человеческих знаний. Но, в отличие от Луллия, Лейбниц считал необходимым сверх того создать для них целесообразную систему обозначений – универсальную характеристику, а также выработать систему правил для получения сложных понятий из простейших. В позднейших философских сочинениях Лейбниц даже заявляет, что Бог творит путем исчисления возможных миров и выбора наилучшего, т. е. фактически по методике Луллия.

В начале XVII века с обоснованием идеи создания всемирного философского языка выступает Р. Декарт, считающий, что в основе такого языка должна лежать классификация всех объектов человеческого мышления.



В 1661 году Джордж Дальгарно издал трактат, посвящённый философскому (или универсальному) языку. По существу Д. Дальгарно разработал словарь основанный на логической классификации понятий, представляющий собой систему философского языка, построенного с помощью условных обозначений категорий и соотношений посредством букв.



Всего Д. Дальгарно выделяет 17 классов, закрепляя за каждым определенную согласную букву (например, n – конкретные предметы, k – отношения и т. п.). Каждый класс разделяется на подклассы, которые в свою очередь обозначаются второй буквой, гласной (например, ka – служебные отношения, ki – партийные отношения, ki – отношения вражды). В результате подобного разбиения получаются такие, например, слова: Нука – слон, Нуkn – лошадь, Нуке – осел.

THE WORKS OF
GEORGE DALGARNO
OF ABERDEEN



REPRINTED AT EDINBURGH.

M.DCCC.XXXIV.

© George C. Marshall Foundation

Сразу же после выхода в свет книги Дальгарно она подверглась критике, что неправильно классифицировала научное знание.

Выдающийся учёный того времени, епископ Джон Уилкинс, попытался исправить недостаток подхода Дальгарно, опираясь на гораздо больший объём сведений.

Джон Уилкинс в 1668 году издаёт трактат, в котором он предложил универсальный язык и десятичную систему мер, которая впоследствии стала основой метрической системы.

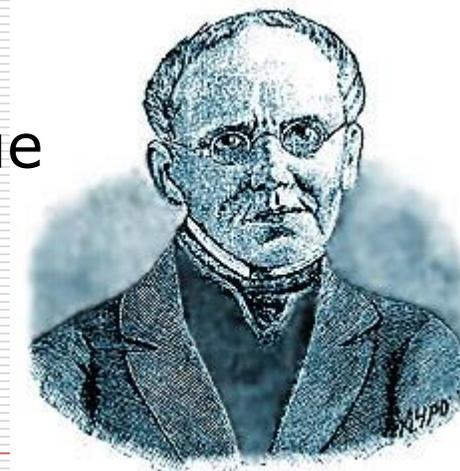


Все понятия, охватываемые языком, Д. Уилкинс делит на шесть типов: трансцендентальные понятия, субстанции, количества, качества, движения, отношения. Затем эти шесть типов в свою очередь подразделяются на 40 классов, каждый из которых обозначается определенным слогом: Ва, Ва, Ве, Ви; Да, Да, Де, Ди; Га, Га, Ге, Ги; и т. п. Присоединение к этим слогам согласных b, d, g, p, t, c, z, s, n приводит к получению определенного числа родов. Указанные согласные, будучи показателями родов, имеют строго определенные значения: b – первый род, d – второй род, g – третий род и т. п.

Дальнейшая спецификация выражается снова присоединением гласных. Таким образом, если согласно таблице классификации понятий стихия на языке Уилкинса будет De, то первый род стихии – огонь, будет Deb, а первый вид огня – пламя будет Deba. Если Ti означает класс ощущаемого качества, то Tid есть второй род этого класса – цвет, а Tida – второй вид, т. е. красный цвет.

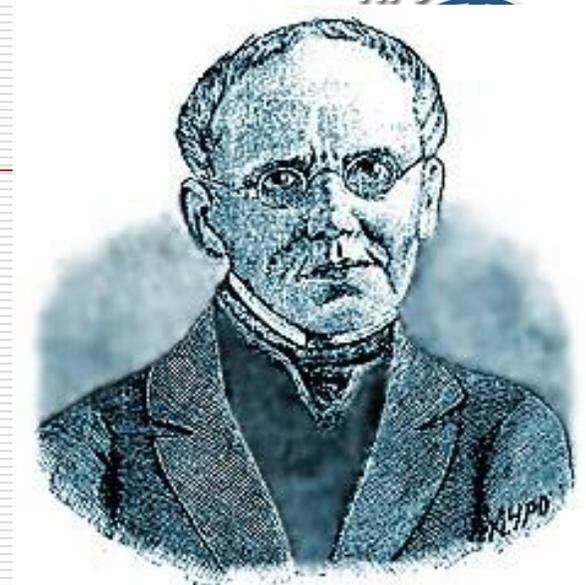
- Следует подчеркнуть, что первые тезаурусы составлялись без всякой связи с особенностями информационной деятельности, они были органически связаны с фундаментальными проблемами познания, отображая представление о мире в целом и закономерностях его постижения средствами естественного языка.

- Коллежский советник Семён Николаевич Корсаков (1787—1853) ставил задачу усиления возможностей разума.
- В 1832 году С. Н. Корсаков опубликовал описание пяти изобретённых им механических устройств, так называемых «интеллектуальных машин», для частичной механизации умственной деятельности в задачах поиска, сравнения и классификации.
- В конструкции своих машин Корсаков впервые в истории информатики применил перфорированные карты.



Машина Корсакова

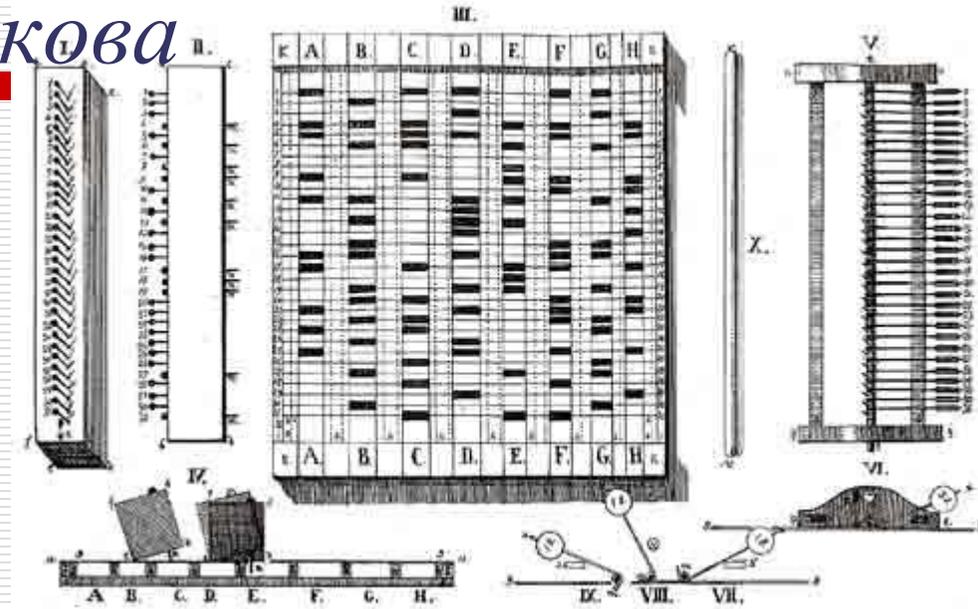
В первой половине XIX века Семён Николаевич Корсаков выдвинул концепцию усиления возможностей разума посредством разработки научных методов и устройств. В 1832 году он опубликовал описание изобретенных им механических устройств, так называемых «интеллектуальных машин».



В своих машинах Корсаков впервые предложил использовать перфорированные карты для задач информационного поиска и классификации. В работах Корсакова содержится целая плеяда новых для того времени идей, как то: многокритериальный поиск с учетом относительной степени важности различных критериев (весовых коэффициентов), способ обработки больших массивов данных – предтеча современных экспертных систем – попытка определить понятие алгоритма.

Машина Корсакова

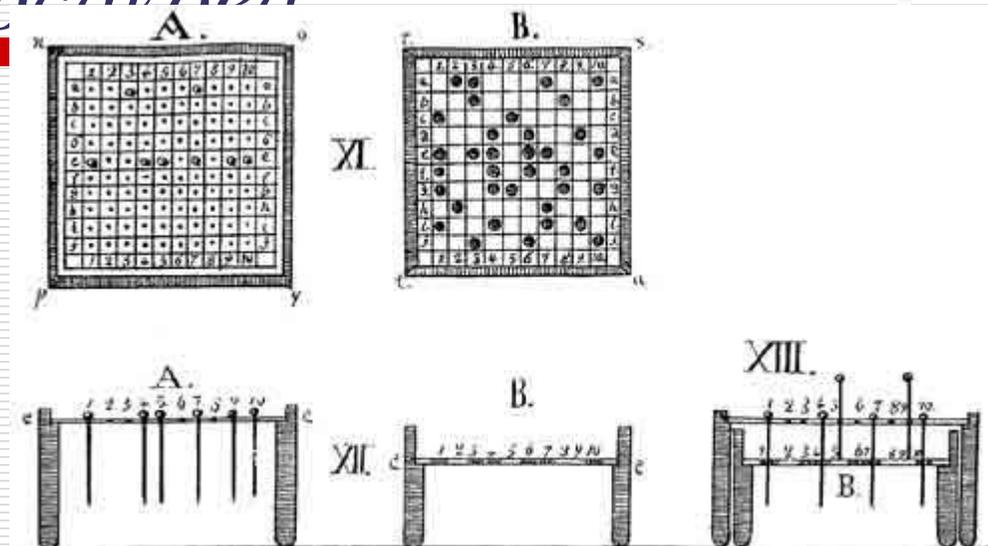
Гомеоскопы прямолинейные:
 прямолинейный с
 неподвижными частями;
 прямолинейный с подвижными
 частями



Пользуясь им можно найти среди большого числа записей, отображённых в гомеоскопической перфорированной таблице, ту, которая содержит все признаки другой заданной записи. Гомеоскоп с подвижными частями может указывать то же самое, что и гомеоскоп прямолинейный с неподвижными частями, и в дополнение к этому он находит и отделяет из заданной записи все те признаки, которые соответствуют (или не соответствуют) аналогичным признакам других записей в таблице.

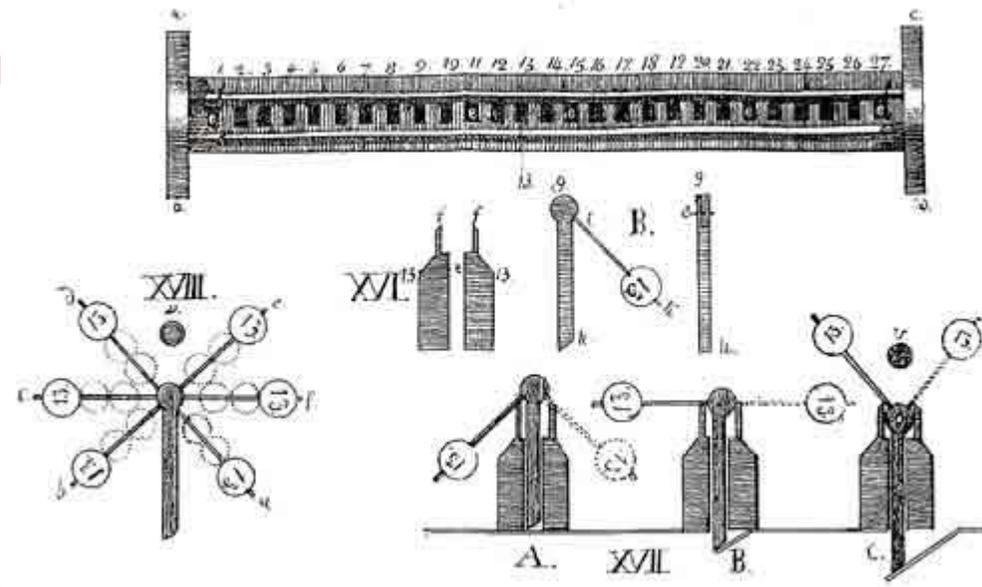
Машина Корсакова

Гомеоскоп плоский



Плоский гомеоскоп аналогично указывает соответствия, имеющиеся у сравниваемых между собой записей, число признаков которых может достигать многих тысяч. С. Н. Корсаков утверждает, что число признаков можно довести до одного миллиона, используя, так называемые, градуированные стержни. В целом плоский гомеоскоп позиционируется Корсаковым как устройство для обработки больших массивов данных.

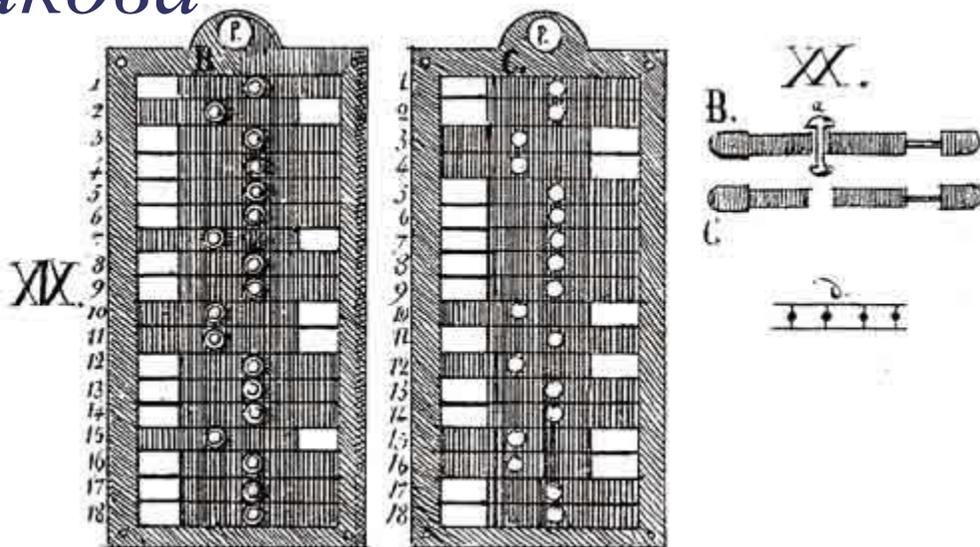
Идеоскоп



Идеоскоп одновременно позволяет выполнить исчисление следующих значений: множество вообще возможных признаков, но отсутствующих в заданной и сравниваемой записях; множество признаков заданной записи, но которых нет в сравниваемой записи из идеоскопической таблицы; множество общих признаков для заданной и сравниваемой записей; множество общих наиболее важных признаков; множество наиболее важных признаков сравниваемой записи из таблицы, но которые отсутствуют в заданной записи; множество признаков сравниваемой записи из таблицы, которые отсутствуют в заданной записи.

Машина Корсакова

Простой компаратор



Компаратор определяет те же операции с множествами, что и идеоскоп. Преимущество компаратора заключается в том, что признаки сравниваемых идей можно задать непосредственно (динамически) перед началом сравнения, не требуется заранее подготавливать и использовать перфорированные таблицы. Ограничение состоит в том, что за один раз возможно сравнение только двух идей.

- В работах Корсакова содержится целый ряд новых для того времени решений, как то:
 - многокритериальный поиск с учетом относительной степени важности различных критериев (весовых коэффициентов),
 - способ обработки больших массивов данных,
 - предтеча современных экспертных систем,
 - попытка определить понятие алгоритма.
- ❖ Он предложил общий принцип поиска или сравнения идей (записей, фактов) на основе их деталей (признаков).

- Таким образом, Корсаков, выражаясь современным языком, определяет информационную запись, набором ее признаков – координатное индексирование по Солтону.
- Словарь Корсакова представлял перфорированные таблицы, где каждый столбец определяет некоторую идею, а в строках перфорируются признаки этой идеи. Перфорация отверстий обеспечивает возможность механического поиска и сравнения идей на основе их признаков.

- Исторически тезаурусы создавались для ручного индексирования документов и при их создании не принимались во внимание вопросы, связанные с автоматической индексацией.
- Трудность построения тезауруса, соответствующего всему тематическому многообразию индексируемой информации, является основной причиной его непопулярности в современных информационных системах.
- Но эффективность информационно-поисковых систем для поддержки научно-образовательной деятельности напрямую зависит от использования специализированных тезаурусов.

- Один из первых в истории и наиболее известных на сегодня тезаурусов (идеографических словарей) составлен британским лексикографом Питером Марком Роже и опубликован в 1852 г. тезаурус Роже (Roget's Thesaurus) — оригинальное название Thesaurus of English Words and Phrases – «Тезаурус английских слов и фраз». Непосредственным и предшественниками словаря П. М. Роже были уже упоминавшиеся работы Д. Далгарно и Д. Уилкинса, санскритский словарь «Амаракоша», переведенный на английский язык в 1808 году.

- Роже в полной мере использовал опыт своих предшественников, он пишет: «Принцип, которым я руководствовался классифицируя слова, является тем же самым, который используется при классификации особей в различных областях естественной истории. Поэтому разделы, выделенные мной, соответствуют естественным семьям ботаники и зоологии, а ряды слов сцементированы теми же отношениями, которые объединяют естественные ряды растений и животных».

- В предисловии к своему словарю он пишет:
....Какой бы живостью ни обладало наше воображение, как бы ни переполняли нас чувства, мы часто попадаем в такое положение, когда нам не хватает слов, чтобы точно выразить свою мысль. Единственно необходимое слово зачастую бежит нашей памяти и мы вынуждены обходиться словами слишком сильными или слишком слабыми, слишком общими или излишне конкретными. Помощь, которую оказывает этот словарь, состоит в предоставлении богатейшего набора слов и выражений, исчерпывающих все оттенки и нюансы каждой общей идеи..

□ Исходя из общих методологических принципов классификации понятий можно говорить о трех типах идеографических словарей:

1. Идеографический словарь-тезаурус — основан на логической рубрикации всего понятийного содержания лексики. Главной задачей является идентификация и последующая рациональная классификация понятийных групп, реально представленных в лексике языка, характер и количество лексико-семантических групп определяется смысловой емкостью языка.

2. Аналогический словарь-тезаурус — основан на выделении тематических лексико-семантических групп, которые располагаются в порядке алфавитного следования тематических доминант (слов-центров).
3. Тематический (учебный) словарь-тезаурус — главной задачей выбор .тем., характеризующий выбранную предметную область (актуальных для учебных целей).

Основные требования, которые предъявляются к классификационной системе, сводятся к следующему:

- а) классифицироваться должны не слова определенного языка, а понятия, что обеспечивает универсальность системы;
- б) классифицируются понятия исходные, лежащие в основе языка;
- в) направляющим принципом классификации является осознание системы понятий как определенным образом организованного единства, расчленение которого должно вестись в естественной последовательности.

- Поль Отле - 1905, 1934
- Эмануэль Гольдберг – 1939
- У. Баттен (Wilfred Batten) – 1939 (к.и.)
- Ванивер Буш – (1939) 1945
- Кельвин Муэрс – 1947
- М. Таубе – 1953
- М. Мастерман в 1954
- А.Кент – 1956 машинный поиск
- Д. Прайс – 1963
- Дж.Солтон – 1975

- А.И.Михайлов - 1976
- В.А.Успенский – 1963
- Ю.А.Шрейдер – 1980 - тезаурус как система знаний, отраженных языком, тогда тезаурус становится интересным сам по себе, а не только как вспомогательный инструмент.
- А.И.Черный – 1975
- Р.С.Гиляревский – 1975
- В.К.Финн – 1975

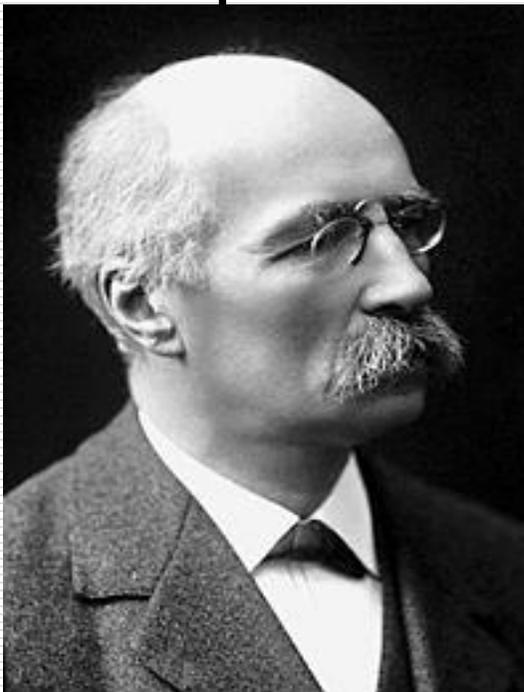
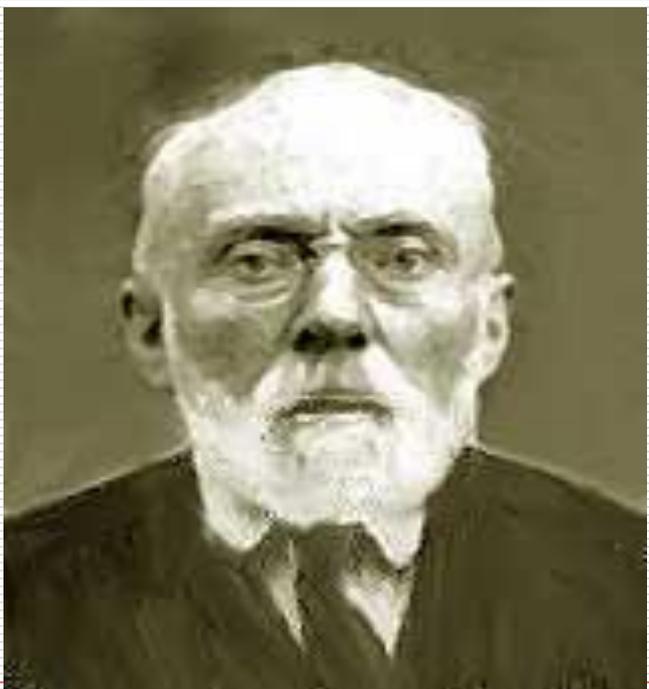
- Питер Марк Роже – 1852

Qui scit ubi scientia habenti est
proximus.

Тот, кто знает, где найти знание,
ближе всего к тому, чтобы знать.

Поль Отле

Поль Отле и Анри Лафонтен *Traité de documentation*



- «Интернет» — по крайней мере как концепция информационной вселенной — был изобретен почти 80 лет назад бельгийским мыслителем Полем Отле, как развитие телефонной связи и телевидения.

Благодаря этим инструментам повсеместности, универсальности и вечности человек приблизится, таким образом, к положению божества...

Поль Отле

Traité de documentation

Ещё в 1934 году Поль Отле, что провода и радиоволны, объединяющие мир, в один прекрасный день будут служить не только для болтовни и развлечений. Они принесут знания в каждый дом.

Мечтатель описал систему, названную им Radiated Library: каждый сможет позвонить в библиотеку и получить информацию по нужной теме (конечно, ввести запрос в поисковую систему в Интернете намного проще).

Так вот, библиотекари разыщут соответствующие книги и отправят изображения страниц на экран телевизора. Эту услугу Отле назвал Televised Book. Он отмечал также, что было бы удобно разделить экран на несколько частей, чтобы иметь возможность сравнивать сразу несколько книг.

Отле осмелился зайти настолько далеко, что предположил полное исчезновение традиционных книг.

Самое интересное, что мы не придумали ничего нового с тех пор, как Отле предложил скрестить телефон и телевизор.

«Мунданеум» — библиотека будущего в представлении Поля Отле.

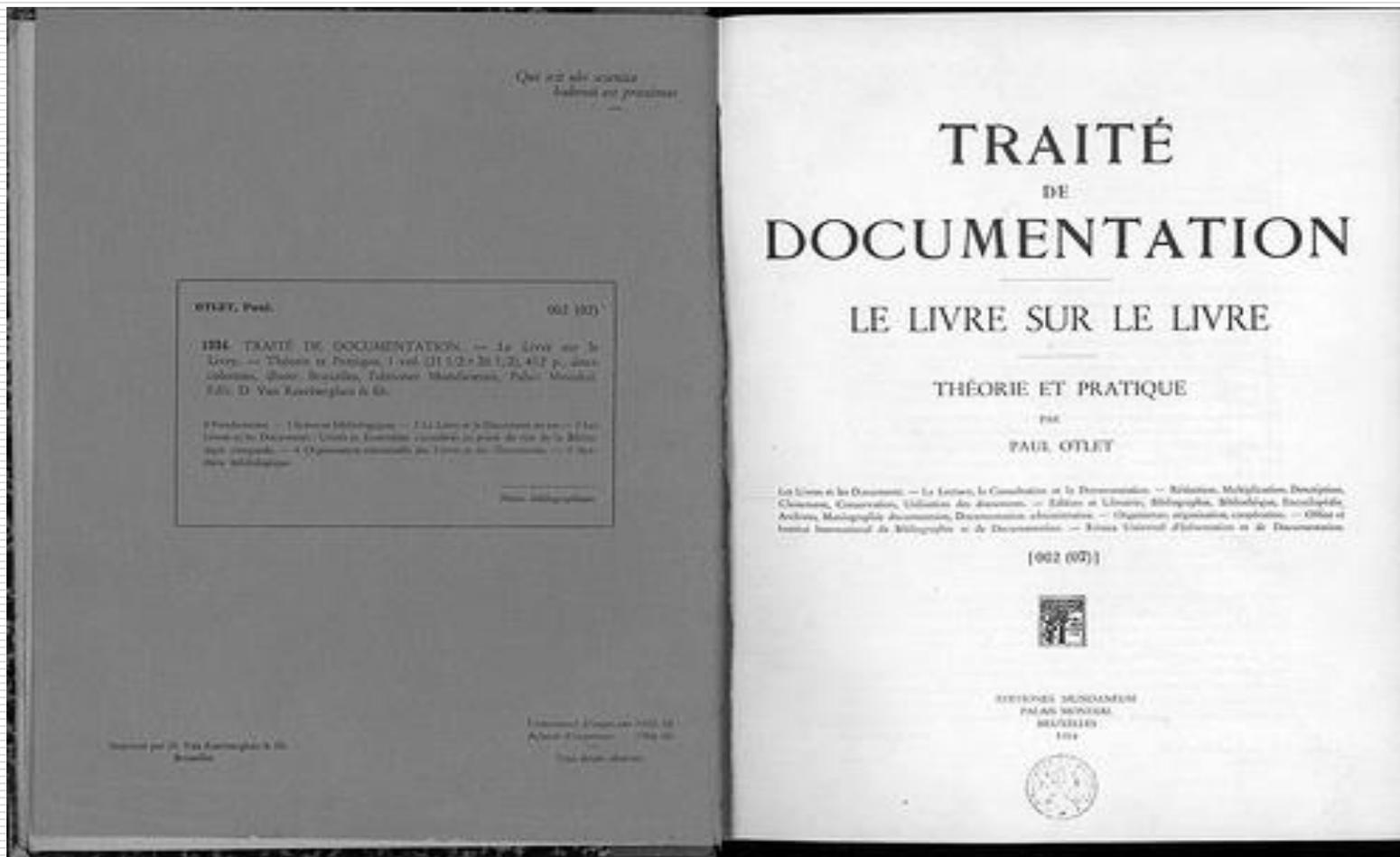


" ...рабочий стол не загружен более никакими книгами. Вместо них возвышается экран и находится в досягаемости телефон. Там далеко, в огромном здании – все книги и все сведения... Оттуда показывают на экран страницу, которую надо прочитать, чтобы узнать ответы на вопросы, заданные по телефону или беспроводному телефону.

Экран был бы двойным, вчетверо большим или десятикратно, если речь зашла бы об увеличении текстов и документов, которые надо одновременно сопоставить; был бы также подвешен громкоговоритель, если прочитанные данные должны были бы быть дополнены услышанными, если видение должно было бы быть дополнено слушанием. ... Утопия сегодня, потому что она ещё нигде не существует, она могла бы оказаться действительно реальностью завтра, лишь бы только ещё усовершенствовались наши методы и инструменты. И это усовершенствование могло бы дойти, возможно, до того, чтобы сделать автоматическим вызов документа на экран ..., а также последовательное их проецирование..."

Поль Отле. Трактат о документации. 1934

Поль Отле. Трактат о документации. 1934



Цели Документации состоят в том, чтобы суметь предложить документированные ответы на запросы по любому предмету в любой области знания:

- 1) универсальные по содержанию;
- 2) точные и истинные;
- 3) полные;
- 4) оперативные;
- 5) отражающие последние данные;
- 6) доступные;
- 7) заранее собранные и готовые к передаче;
- 8) предоставленные как можно большему числу людей



Поль Отле о будущем книги



«Мунданеум» начал работать в 1910 году (Поль Отле стоит в центре), но подготовка к нему началась ещё в 1895-м. Он должен был стать систематическим архивом знаний всего человечества.



Мунданеум

Когда нацистская Германия вторглась в Бельгию в 1940 году, в Mundaneum была размещена выставка искусства Третьего рейха, и большая часть материалов и оборудование были потеряны.



Fig. 1. — Projection sur la table de travail.



Fig. 2. — Film serté dans son cadre réduction aux 2/3.

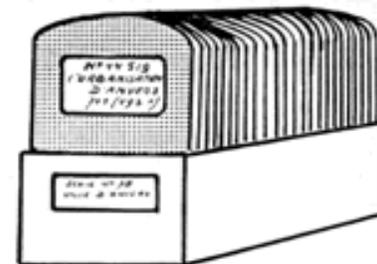


Fig. 3. — Série de cadres.

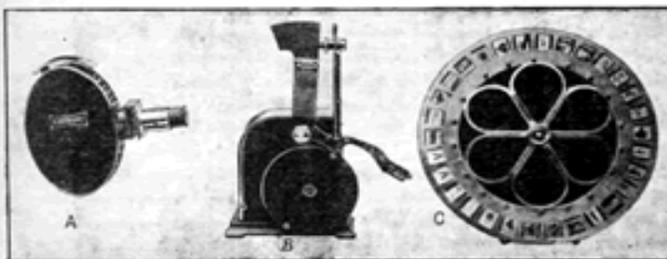


Fig. 4. — Lampe et disque recevant les films en cadre.

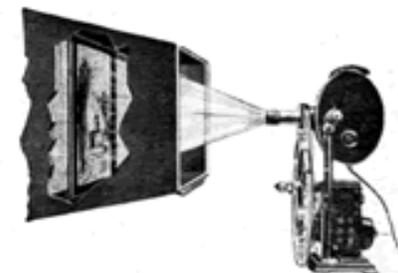


Fig. 5. — Projection en plein jour dans une caisse d'ombre.



Fig. 6. — Agrandissement de la photographie sur film d'un article extrait d'une revue.

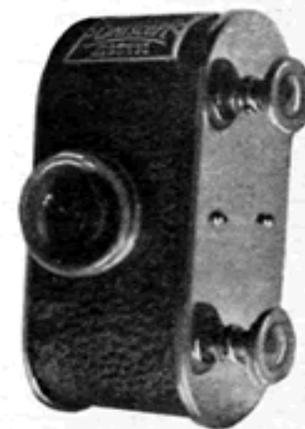


Fig. 8. Loupe à dérouleur pour film.



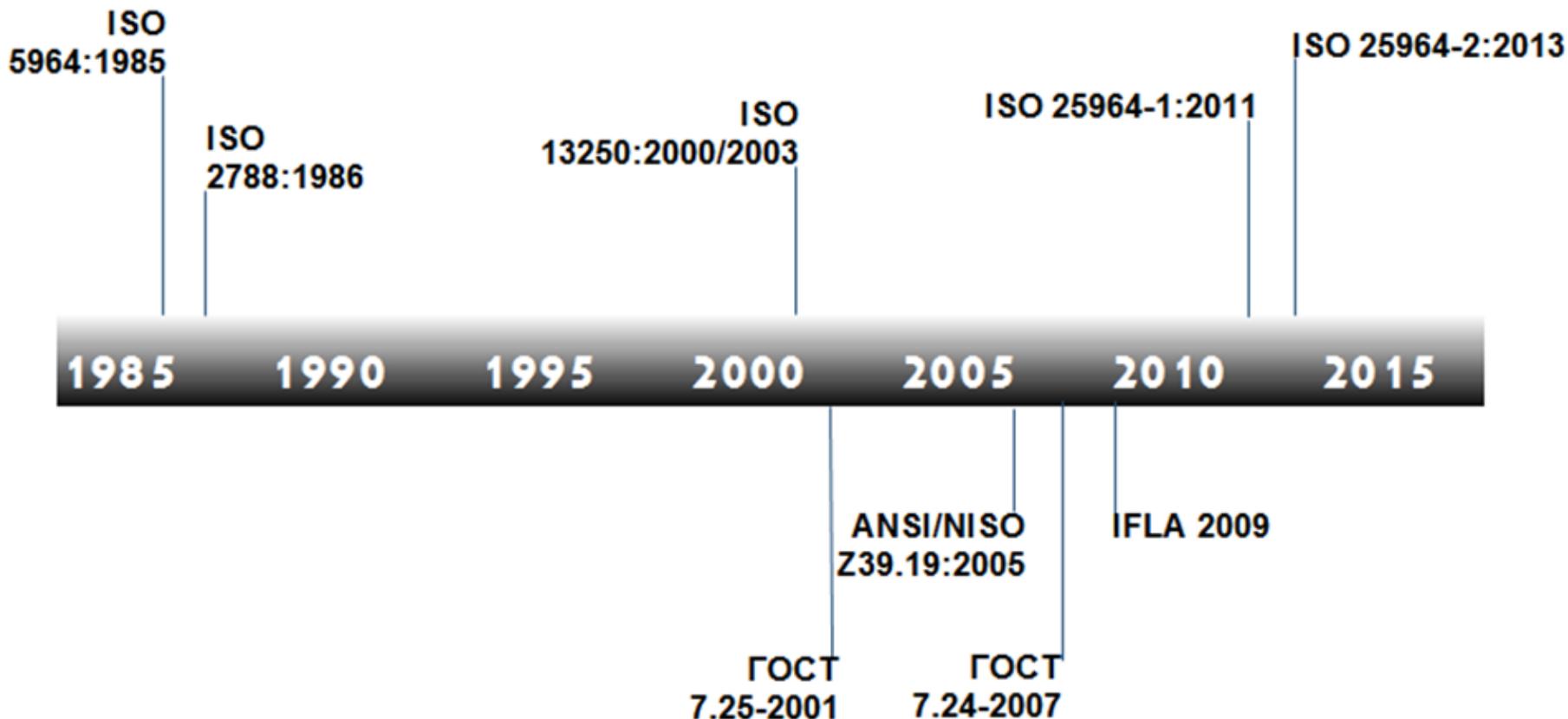
En 1934, le fondateur du Mundaneum
Paul Otlet imaginait l'Internet

- Информационно-поисковый тезаурус (в соответствии с определениями стандартов) – это нормативный (контролируемый) словарь ключевых терминов на естественном языке с явно указанными семантическими отношениями между терминами, предназначенный для описания содержания документов и поисковых запросов.
- Тезаурус предназначен для описания конкретной предметной области, каждый термин которого обозначает или описывает какое-либо понятие из данной предметной области.

- Отметим, что описанная в национальных и международных стандартах модель информационно-поискового тезауруса предназначена для его использования в процессе ручного, экспертного анализа документов.
- Основной целью разработки традиционных информационно-поисковых тезаурусов является использование их единиц (дескрипторов) для описания основных тем документов в процессе ручного индексирования.

- Поэтому важно, чтобы набор дескрипторов информационно-поискового тезауруса позволял описывать тематику документов предметной области.
- При этом сам процесс индексирования по такому тезаурусу базируется на лингвистических, грамматических знаниях, а также знаниях о предметной области, которые имеются у профессиональных индексаторов текстов.

- Индексатор сначала должен прочитать текст, понять его и затем изложить содержание текста, пользуясь дескрипторами, указанными в информационно-поисковом тезаурусе.
- Индексатор должен хорошо понимать всю терминологию, использованную в тексте, – для описания основной темы текста ему понадобится значительно меньшее количество терминов.



- Стандарт NISO Z39.19:2005 предполагает, что контролируемые словари, обычно используемые для описания содержания и определения понятий, а также используются для представления метаданных ассоциируемых с содержанием объектов.
- В отличие от других стандартов этот стандарт подразумевал также использование информатизированных источников в схемах метаданных, таких как Dublin Core и стандартных сетевых протоколов доступа, таких как протокол Z39.50.

- Британский стандарт BS8723:2008 (Структурированные словари для поиска информации). Он состоит из пяти частей, в котором последняя посвящена протоколам и формату обмена для взаимодействия.
- Она направлена на структурированные словари в целом, характеризуя и сравнивая разные типы KOS (Knowledge Organization System) такие как классификация схем, таксономия, схемы оглавления материала, тезаурус, списки авторов и онтология.

- В новом Стандарте ISO 25964-1:2011, ISO 25964-2:2013 (*Информация и документация. Тезаурус и взаимодействие с другими словарями. Тезаурус для поиска информации*) были пересмотрены требования к процессам индексации, а также применению сетевых протоколов доступа.

- Наиболее значительными дополнениями были американский стандарт NISO Z39.19:2005 и британский BS5723-2008.
- NISO обобщал стандарт для «контролируемых словарей».
- Британский стандарт устанавливал связь между тезаурусом и тем, что не является тезаурусом с контролируемым словарем.

- Наиболее существенное отличие состоит в переходе от терминологической структуризации к концептуальному представлению, в котором эквиваленты лежат между понятиями, а не между словами.

- Впервые в стандарте ISO 25964-1:2011 представлены модели (схемы) данных для сетевого взаимодействия. В основу схем данных положены модель данных протокола Z39.50 (схема данных Zthes) и рекомендации SKOS (схема данных SKOS), базирующейся на RDF модели представления понятий.
- Simple Knowledge Organization System (SKOS) – рекомендации W3C.

- Как приложение RDF SKOS позволяет публиковать термины в веб среде, связывать их с информационными элементами, а также включать их в другие концептуальные схемы.
- Zthes шире по набору атрибутов SKOS и позволяет делать все тоже самое, но только через XSLT-преобразование.

- Существуют и другие схемы данных для представления тезаурусов и контролируемых словарей, но все они беднее по описанию свойств и не соответствуют стандарту.
- Можно отметить следующие схемы: MARC, VDEX и MODS (предназначенных в основном для управляемых словарей).

- Для реализации тезауруса наиболее подходит схема данных Zthes, как самая продвинутая из стандартных схем.
- Главное преимущество схемы Zthes состоит в том, что она соответствует модели сетевого протокола Z39.50, что позволит работать не только собственном локальным тезаурусом, а подключать при необходимости тезаурусы, расположенные в сети.

Полисемия

Словарные статьи в коллекции: (public_cat = Thesaurus of Information Technology: Dictionary Articles)

информация

Информация - сведения (сообщения, данные) независимо от формы их представления (Федеральный закон «[Об информации, информатизации и защите информации](#)»).

Информация (в процессах ее обработки) - любой факт, понятие или значение, полученные из данных, а также контекст, выбранный из знаний, или контекст, ассоциированный со знаниями (стандарт [ISO/IEC 2382-1](#)).

Информация - сведения, воспринимаемые человеком и (или) специальными устройствами как отражение фактов материального или духовного мира в процессе коммуникации (ГОСТ 7.0-99. Информационно-библиотечная деятельность, библиография).

В настоящий момент в литературе нет четкого определения, что такое "информация" - считается, что это понятие является интуитивно понятным, как понятие точки в геометрии. В обиходе "информацией" называют любые данные или сведения, которые кого-либо интересуют. Например, сообщение о каких-либо событиях, о чьей-либо деятельности и т.п. "Информировать" в этом смысле означает "сообщить нечто, неизвестное раньше". С середины 20 века понятие "информация" стало общенаучной категорией, что было связано с всеобщим распространением ЭВМ и становлением науки "информатика", а так же введением количественной меры информации, разработкой количественной теории информации (по [Шеннону](#) и [Колмогорову](#)).

Ниже приводятся некоторые определения понятия информация:

1. **ИНФОРМАЦИЯ** (от лат. informatio - разъяснение, осведомление) - любые сведения и данные, отражающие свойства объектов в природных (биологических, физических и др.), социальных и технических системах и передаваемые звуковым, графическим (в том числе письменным) или иным способом без применения или с применением технических средств.
2. **ИНФОРМАЦИЯ** — это обозначение содержания сообщения, полученного из внешнего мира в процессе нашего приспособления к нему и приспособления к нему наших чувств.
3. **ИНФОРМАЦИЯ** - (по Колмогорову) «Информация есть характеристика не сообщения, а соотношения между сообщением и его потребителем. Без наличия потребителя, хотя бы потенциального, говорить об информации бессмысленно».
4. **ИНФОРМАЦИЯ** - часть знаний, которая используется для ориентирования, активного действия, управления, т.е. в целях сохранения, совершенствования, развития системы (Н. Винер).

Важность (ценность) какой-либо информации зависит от множества обстоятельств и, по существу, не поддается формализации. Однако можно четко определить основные свойства информации исходя из подхода к ее определению. Условно

Информатика

Информатика - в широком смысле - отрасль знаний, изучающая общие свойства и структуру информации (данных и знаний), а также закономерности и принципы ее создания, преобразования, накопления, передачи и использования в различных областях человеческой деятельности.

Информатика - в узком смысле - отрасль знаний, изучающая законы и методы накопления, передачи и обработки информации с помощью компьютера.

Функция информатики заключается в разработке методов и средств преобразования информации и организации технологического процесса переработки информации.

Задачи информатики: исследование информационных процессов любой природы; разработка информационной техники и создание новейшей технологии переработки информации на базе полученных результатов исследования информационных процессов; решение научных и инженерных проблем создания, внедрения и обеспечения эффективного использования вычислительной техники во всех сферах общественной жизни.

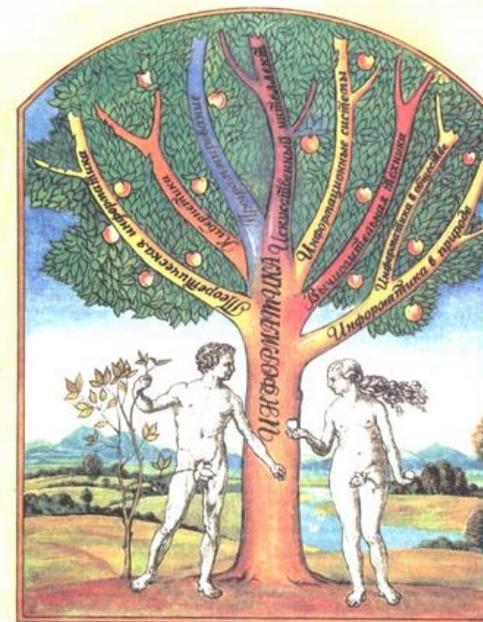
Ключевые термины, связанные с термином : "Информатика":

1. [Доцифровая информатика \[ru\]](#)
2. [Информатизация \[ru\]](#)
3. [Информационная революция \[ru\]](#)
4. [Информационная технология \[ru\]](#)
5. [информация \[ru\]](#)
6. [История информатики \(ИИТ\) \[ru\]](#)
7. [Модель \[ru\]](#)

Литература

Основная

1. [Кнут Д. Программирование как искусство // В кн.: Из истории кибернетики. Редактор-составитель Я.П. Фет. – Новосибирск: Академическое издательство «Гео». - 2006. - С. 95-109.](#)
2. [Лаеров С. С. Что изучает информатика? // Компьютерные инструменты в образовании. - 2000. - Выпуск № 1. - С. 13-17.](#)
3. [Поспелов Л. А. Становление информатики в России // В кн.: «Очерки истории информатики в России». Составители – Л. А.](#)



Дерево информатики (Д.А.Поспелов, 1994)



Тезаурус [ru]

Тезаурус - совокупность терминов, описывающих данную предметную область, с указанием семантических отношений (связей) между ними.

– это словарь, в котором слова и словосочетания с близкими значениями сгруппированы в единицы, называемые понятиями, концептами или дескрипторами, и в котором явно (в виде отношений, иерархии) указываются семантические отношения между этими понятиями (концептами, дескрипторами).

(NT) Тезаурус (add) [ru]

Головные термины:

[BT] [Система организации знаний \[ru\]](#)

Дочерние термины:

[NT] [Ақпараттық-іздеу тезаурусы \[kz\]](#)

[NT] [Информационно-поисковый тезаурус \[ru\]](#)

[NT] [Лексическая единица ИПЯ \[ru\]](#)

[NT] [Тезаурус WordNet \[ru\]](#)

[NT] [Тезаурус Роже \[ru\]](#)

[NT] [Тезаурус РуТез \[ru\]](#)

[NT] [Тезаурус ЮНЕСКО \[ru\]](#)

[NT] [Управляемый словарь \[ru\]](#)

Ключевые термины, связанные с термином

[RT] [База знаний \[ru\]](#)

[RT] [Дескрипторный словарь \[ru\]](#)

[RT] [Иерархический указатель информационно-поискового тезауруса \[ru\]](#)

[RT] [Искусственный интеллект \[ru\]](#)

[RT] [Лексико-семантический указатель информационно-поискового тезауруса \[ru\]](#)

[RT] [Лексическая единица \[ru\]](#)

[RT] [Многоязычный информационно-поисковый тезаурус \[ru\]](#)

[RT] [Одноязычный информационно-поисковый тезаурус \[ru\]](#)

[RT] [Онтология \[ru\]](#)

[RT] [Парадигматические отношения \[ru\]](#)

[RT] [Пермутационный указатель информационно-поискового тезауруса \[ru\]](#)

Словарные статьи в коллекции: (public_cat = Thesaurus of Information Technology: Dictionary Articles)

Тезаурус

Тезаурус (Thesaurus) (от греческого «θησαυρος» (thesauros) – сокровище) — в обобщенном смысле это словарь ключевых терминов (понятий), описывающих некоторую предметную область, с указанием семантических отношений (связей) между терминами (понятиями).

Тезаурусы, прежде всего, используются для систематизации, классификации и поиска информационных ресурсов. Заметим, что кроме классификации ресурсов, тезаурус может использоваться как самостоятельная база знаний, показывая место тех или иных понятий в предметной области.

В литературе имеются большое количество определений тезауруса, уточняющих приведенное определение применительно к конкретной предметной области, которые функционально эквивалентны ([толерантны](#)) приведенному. Приведем некоторые из них:

1. Тезаурус - это полный систематизированный набор данных (понятий) о какой-нибудь области знаний, который позволяет человеку или вычислительной машине в ней ориентироваться.
2. Тезаурус - это совокупность терминов, описывающих данную предметную область, с указанием семантических отношений (связей) между ними.
3. Тезаурус - это словарь с дополнительной информацией о (семантических) связях терминов, таких как синонимы, омонимы, родо-видовые отношения, часть/целое.
4. Тезаурус - это упорядоченный управляемый словарь, структурированный таким образом, что в нем между терминами четко определены и идентифицированы отношения эквивалентности, гомографии, иерархии и ассоциации с использованием стандартизированных индикаторов этих отношений.
5. Тезаурус - это словарь, специальная терминология, собрание сведений, корпус или свод, полномерно охватывающие понятия, определения и термины конкретной области знаний или сферы деятельности.
6. Тезаурус - в современной лингвистике - это особая разновидность словарей общей или специальной лексики, в которых указаны семантические отношения (синонимы, антонимы, паронимы, гипонимы, гиперонимы и т. п.) между

Тезаурус Роже

Тезаурус Роже (Roget's Thesaurus, оригинальное название Thesaurus of English Words and Phrases — «Тезаурус английских слов и фраз») — один из первых в истории и наиболее известных на сегодня идеографических словарей. Составлен британским лексикографом Питером Марком Роже и опубликован в 1852 г.

В словарях - тезаурусах лексика организуется по тематическому принципу. Тезаурус Роже включает шесть основных семантических категорий:

1. Абстрактные отношения;
2. Пространство;
3. Материя;
4. Разум;
5. Воля;
6. Чувственные и моральные силы.

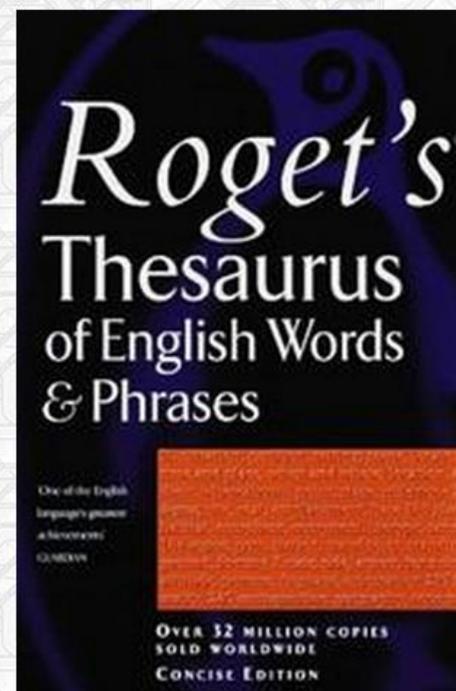
Эти классы подразделяются на 24 подкласса. Например, в классе пространство выделяются 4 подкласса:

- 2.1 Пространство вообще;
- 2.2. Измерение;
- 2.3. Форма;
- 2.4. Движение.

В рамках этих подклассов выделяются более мелкие подклассы и т.д., вплоть до самых мелких, далее не подвергающихся членению тематических групп, которых насчитывается около тысячи.

Однако вскоре после опубликования словаря стало ясно, что пользоваться им очень трудно. Практика потребовала разумного синтеза идеографического и алфавитного расположения слов. Поэтому Роже приложил к словарю алфавитный указатель, снабдив каждое слово информацией о его месте в идеографической классификации. Тезаурус Роже следует признать выдающимся явлением в мировой лексикографии. Его главное достоинство заключается в том, что это была первая научно обоснованная попытка создать некоторый макет логически упорядоченной лексики языка.

Главными недостатками тезауруса Роже является чрезмерная схематичность при отнесении слов в рубрики – стремясь по максимуму воспроизвести логические отношения слов языка, главным образом родо-видовые, Роже упускал из виду множество других отношений, которыми связаны слова, например, «пространство» выделено в отдельную категорию, а «время» входит в категорию «абстрактные понятия», слово музыка входит в группу «звук» и далее – в большую категорию «материя», а живопись –



18 января 1779 г., Сохо, Лондон, Великобритания - 12 сентября 1869 г., Западный Малверн, Великобритания



Основные достижения:

Питер Марк Роже (Роджет) (Peter Mark Roget) - британский врач, физик, богослов, лексикограф - составитель одного из первых в истории и наиболее известных на сегодня [идеографических словарей Тезаурус Роже](#) (англ. Roget's Thesaurus, оригинальное название Thesaurus of English Words and Phrases — «Тезаурус английских слов и фраз», опубликован в 1852 г.).

9 декабря 1824 года Роже представил доклад, озаглавленный "Объяснение оптического обмана возникающего при вращении спиц в колесе. Исследования Питера Марка Роджета, опубликованные в этой статье привели к открытию кинематографа. На основе этих работ английский физик Фарадей сконструировал в 1830 г. прибор, названный по его имени «колесом Фарадея» и описанный во всех трудах по физике, а Джон Гершель, поставив новый опыт занимательной физики, создав первую оптическую игрушку, в которой использовались рисунки.

Краткая биография:

Питер Марк Роже родился в Лондоне в семье швейцарского священника, учился медицине в Университете Эдинбурга, который окончил в 1798 году. Роже ушел из профессиональной жизни в 1840 году, в 1848, начал подготовку к изданию работы, которая должна была

увековечить его память. Это было Каталог слов, организованных по смыслу. Его первое печатное издание вышло в 1852 году под названием "Тезаурус английских слов и фраз". В этом каталоге слова классифицируются и упорядочиваются таким образом, "чтобы содействовать выражению идей и оказанию помощи в литературной композиции". В течение его жизни работа переиздавалась двадцать восемь раз; после его смерти была пересмотрена и расширена его сыном, Джоном Льюисом Роже (1828-1908). Роже умер во время отпуска в Западном Малверн в возрасте 90 лет, и был похоронен там же на кладбище церкви Сент-Джеймс.

Ключевые слова: [Тезаурус Роже](#); [Тезаурус](#); [идеографический словарь](#); лексикография; кинематограф;



Архитектура фон Неймана [ru]

Архитектура фон Неймана В настоящее время фон-неймановской архитектурой называется организация ЭВМ, при которой вычислительная машина состоит из двух основных частей — линейно-адресуемой памяти, слова которой хранят команды и элементы данных, и процессора, выполняющего эти команды. В основе модели вычислений фон Неймана лежат принцип последовательной передачи управления (счётчик команд) и концепция переменной (идентификатор). Большинство современных ЭВМ по своей структуре отвечают принципу программного управления. Типичная фон-неймановская ЭВМ содержит: память, устройство управления, арифметико-логическое устройство и устройство ввода/вывода.

(NT) Архитектура фон Неймана (add) [ru]

Синонимы:

[SYN] фон-неймановской архитектура

Головные термины:

[BT] Архитектура вычислительной машины [ru]

Дочерние термины:

[NT] Вычислительная машина с непосредственными связями [ru]

[NT] Гарвардская архитектура [ru]

[NT] Машина Тьюринга-Поста [ru]

[NT] Модель вычислений фон Неймана [ru]

[NT] Модель вычислителя [ru]

[NT] Принципы фон Неймана [ru]

[NT] Управление последовательностью команд [ru]

Ключевые термины, связанные с термином

[RT] Машина Тьюринга-Поста [ru]

● **Контекстный поиск:** Задайте образец для поиска:

Найти

Словарные статьи в коллекции: (public_cat = Thesaurus of Information Technology: Dictionary Articles)

Архитектура фон Неймана

Синонимы: Архитектура фон Неймана; неймановская архитектура; принстонская архитектура; фон-неймановская архитектура; фон неймановская архитектура;

В статье фон Неймана «First Draft of a Report on the EDVAC» определены основные устройства [ЭВМ](#), с помощью которых должны быть реализованы его принципы. Большинство современных ЭВМ по своей структуре отвечают принципу программного управления. Типичная фон-неймановская ЭВМ содержит: память, устройство управления, арифметико-логическое устройство и устройство ввода/вывода.

В настоящее время фон-неймановской архитектурой называется организация ЭВМ, при которой вычислительная машина состоит из двух основных частей — линейно-адресуемой памяти, слова которой хранят команды и элементы данных, и процессора, выполняющего эти команды. В основе модели вычислений фон Неймана лежат принцип последовательной передачи управления (счётчик команд) и концепция переменной (идентификатор).

Ключевые термины, связанные с термином : "Архитектура фон Неймана":

1. [Архитектура вычислительной машины \[ru\]](#)
2. [Вычислительная машина с непосредственными связями \[ru\]](#)
3. [Гарвардская архитектура \[ru\]](#)
4. [Машина Тьюринга-Поста \[ru\]](#)
5. [Модель вычислений фон Неймана \[ru\]](#)
6. [Модель вычислителя \[ru\]](#)
7. [Принципы фон Неймана \[ru\]](#)
8. [Процедурный язык \[ru\]](#)
9. [Управление последовательностью команд \[ru\]](#)

Литература

Основная

1. [Орлов С. А., Цилькер Б. Я. Организация ЭВМ и систем: Учебник для вузов. 2-е изд. — СПб.: Питер, 2011. — 688 с.: ил. - ISBN 978-5-49807-862-5.](#)

Факультативная

1. [John von Neumann. First Draft of a Report on the EDVAC // Journal IEEE Annals of the History of Computing archive \(October 1993\). - 1993. - Volume 15. - Issue No. 4. - P. 27-75.](#)
2. [Нитусов Александр. Первый электронный компьютер \(К 100-летию его создателя — Мориса Уилкса\) // PCWeek. - 2013.](#)
3. [Черняк Леонид. Архитектура фон Неймана. реконфигурируемые компьютерные системы и антимашина // Открытые системы. - 2008. - Выпуск № 6.](#)

Ссылка на персон:

1. [Нейман Джон фон](#)

Ключевые термины статьи: [Архитектура вычислительной машины](#); [Принципы фон Неймана](#); [модель вычислений фон Неймана](#);

Словарные статьи в коллекции: (public_cat = Thesaurus of Information Technology: Dictionary Articles)

Модель вычислений фон Неймана

Модель вычислений фон Неймана основана на [фон неймановской архитектуре ЭВМ](#). В настоящее время фон-неймановской архитектурой называется организация ЭВМ, при которой вычислительная машина состоит из двух основных частей — линейно-адресуемой памяти, слова которой хранят команды и элементы данных, и процессора, выполняющего эти команды.

В основе модели вычислений фон Неймана лежат принцип последовательной передачи управления (счётчик команд) и концепция переменной (идентификатор).

Ключевые термины, связанные с термином : *"Модель вычислений фон Неймана"*:

1. [Архитектура фон Неймана \[ru\]](#)
2. [Императивное программирование \[ru\]](#)
3. [Императивный язык \[ru\]](#)

Ключевые термины статьи: модель вычислений: [Архитектура фон Неймана:](#)

Императивное программирование

Императивное программирование (императивный стиль программирования) — это парадигма программирования, которая описывает процесс работы ЭВМ (**план вычислений**) в виде последовательности инструкций, изменяющих состояние программы. Императивная программа очень похожа на приказы, выражаемые повелительным наклонением в естественных языках, то есть это последовательность команд и логических переходов, которые должен выполнить компьютер.

Основные конструкции:

- Манипулирование ячейками памяти.
- Оператор присваивания.
- Явные операторы передачи управления.
- Циклы, условный оператор.

Императивный стиль программирования основан на [фон неймановской модели вычислений](#). Методология императивного программирования характеризуется принципом последовательного изменения состояния [вычислителя](#) пошаговым образом. При этом управление изменениями полностью определено и полностью контролируемо.

Для решения задачи с использованием императивного стиля программирования необходимо: вначале создать модель вычислений в некоторой формальной системе, а затем записать ее на императивном языке. В качестве математической модели вычислений императивное программирование использует [машину Тьюринга-Поста](#) — абстрактное вычислительное устройство, предложенное на заре компьютерной эры для описания алгоритмов и решения проблемы вычислимости.

Ключевые термины, связанные с термином : "Императивное программирование":

1. [Ассемблер \[ru\]](#)
2. [Императивный язык \[ru\]](#)
3. [Модель вычислений фон Неймана \[ru\]](#)
4. [Программирование структурное \[ru\]](#)
5. [Процедурный язык \[ru\]](#)

Ключевые термины статьи: [Парадигма программирования:](#) [Модель вычислений фон Неймана:](#) [Процедурное программирование:](#) [Процедурный язык:](#) [Императивный язык:](#)

- Два типа информационных потребностей исследователя: потребности в сведениях об источниках необходимой научной информации и потребности в **самой** необходимой научной информации (фактах).
- Для удовлетворения информационных потребностей первого предназначены информационные системы, получившие название «документальных» (библиографических), второго типа – «фактографических».
- В настоящее время наиболее востребованным средством информационного обеспечения научной деятельности становятся интеллектуальные системы, сочетающие возможности информационных систем обоих названных типов и позволяющие удовлетворять информационные потребности квалифицированного пользователя в соответствии со схемой «документ → факт → рассуждение».

- В информационной системе тезаурус является не только самостоятельным информационным ресурсом, но и инструментом для классификации или индексации ресурсов. Пользователь информационной системы должен иметь возможность:
 - Осуществлять просмотр тезауруса.
 - Осуществлять поиск ресурсов по ассоциированным с ними терминам или понятиям.
- Поиск ресурсов может вестись двумя способами:
 - Поиск по ключевым словам, используя тезаурус.
 - Навигация по тезаурусу. То есть поиск сначала нужного понятия в тезаурусе с последующим запросом ресурсов, соответствующих этому понятию.

- При поиске ресурсов по ключевым словам поисковая система может, используя тезаурус, расширять результаты поиска, выдавая пользователю не только ресурсы, соответствующие введенным пользователем ключевым словам, но и ресурсы, соответствующие связанным с ними терминам, или терминам, обозначающим также более узкие понятия относительно исходного термина.
- Например, если пользователь ищет ресурсы, соответствующие термину «туннель», в результатах поиска необходимо выдать также все ресурсы, соответствующие термину «тоннель», поскольку оба они являются разными правильными вариантами написания одного и того же слова.

- Или если ищутся ресурсы, соответствующие понятию *дифференциальные и функциональные уравнения*, имеет смысл включить в результаты поиска также ресурсы, соответствующие рубрике *системы функциональных уравнений и неравенства*.
- Система поиска может также, используя тезаурус, подсказать пользователю, по каким еще словам ему стоит осуществить поиск (например, квазисинонимы, связанные термины, более широкие термины, и т.д.).
- Оба этих варианта использования тезауруса широко применяются, например, в поисковых машинах.

- Интерфейс просмотра тезауруса должен:
 - Показывать все атрибуты данного термина или понятия.
 - Показывать, с какими терминами и понятиями связан данный термин или понятие.
 - Достаточно наглядно показывать пользователю место термина или понятия в иерархии понятий тезауруса.

- Однако существует ряд тезаурусов, основная задача которых не индексация ресурсов, а их классификация. В этом случае основными объектами таких тезаурусов (*классификаторов*) выступают не термины, а понятия (*рубрики*), и, часто, идентифицирующие их уникальные идентификаторы (коды классификации). Отношения в таком тезаурусе – не семантические связи между терминами, а характеризующие логику описываемой предметной области отношения между понятиями (рубриками).
- Структура классификатора соответствует структуре обычного тезауруса, поскольку связи между его рубриками по смыслу те же, что и между терминами тезауруса, и классификатор является его частным случаем.
- Однако при классификации в соответствии с ресурсами ставятся не термины, а обозначаемые ими понятия. Потому в схеме данных информационной системы понятия тезауруса должны быть выделены в самостоятельные объекты. Это означает, что такая схема должна иметь структуру, отличную от вышеописанных стандартов, в которых понятия не выступают отдельными объектами, а есть лишь термины и связи между ними. В то же время, схема должна позволять работать с тезаурусами, описанными в соответствии с этими стандартами, т.е. быть совместима с ними.

Словари (ключевые признаки, ключевые термины) – формируют особый вид метаданных, которые отражают наиболее существенные свойства объекта, их специфика определяется терминологией конкретной предметной области. Необходимо рассматривать различные типы ключевых терминов, а именно:

- ключевые термины в стандартном понимании;
- ключевые термины, описывающие персону;
- ключевые термины, описывающие организацию;
- ключевые термины, описывающие временные периоды;
- ключевые термины, описывающие географические понятия,
- тематические словари-классификаторы, тезаурусы, описания предметной области данной предметной области.

Стандартных словарей, например, УДК, ГРНТИ MSC2010, ORTELIUS, тезаурус ЮНЕСКО для классификации здесь явно недостаточно.